Efficient Electrical Spin Injection and Realization of a spin-LED

B.T. Jonker, Y.D. Park*
Materials Science and Technology Division

B.R. Bennett
Electronics Science and Technology Division

H.-D. Cheong, G. Kioseoglou, and A. Petrou
State University of New York at Buffalo

Electrical spin injection into a semiconductor is a prerequisite for realizing the potential of semiconductor-based spintronic devices. This has been an elusive goal, however, and only modest effects (≤1%) have been obtained. We report highly efficient electrical spin injection from a magnetic contact into a GaAs-based light-emitting diode (LED) heterostructure — a spin-LED — in which the spin injection efficiency exceeds 50%. The samples are grown by molecular beam epitaxy. The semimagnetic semiconductor ZnMnSe serves as a source of spin-polarized electrons that are injected via an applied bias voltage into a GaAs quantum well. Radiative recombination of spin-polarized carriers results in the emission of circularly polarized light. The quantum selection rules relate the optical polarization to the carrier spin polarization, enabling a quantitative measure of spin injection efficiency. The measured circular polarization of the electroluminescence (EL) exceeds 50%, demonstrating that highly efficient spin transport occurs across the ZnMnSe/AlGaAs interface despite the large 0.5% lattice mismatch.

INTRODUCTION

Spin injection and transport in semiconductor heterostructures is a promising avenue to add new spin-dependent functionality to the many attractive properties of semiconductor devices.[1] The seminal proposal by Datta and Das[2] of a spin-polarized field effect transistor, with magnetic source and drain contacts for spin injection and detection, has stimulated a great deal of effort to better understand the behavior of spin-polarized carriers in semiconductor hosts under conditions of dynamic transport. Optical pumping has routinely been used to create spin-polarized carrier populations in semiconductor heterostructures, and has provided tremendous insight into their behavior.[3] Very recently, extraordinarily long spin lifetimes and diffusion lengths have been observed in optically pumped systems.[4,5] Spin diffusion lengths of many microns have been reported in GaAs,[4] for example, illustrating that a spin-polarized mode of operation is certainly feasible for every modern transport device.

It is very desirable to electrically inject spin-polarized carriers via a magnetic contact to increase the potential for practical applications, as originally proposed. This would provide a very simple and direct implementation of spin injection in which the contact area defines the spin source. It would also significantly broaden application horizons. Electrical spin injection has been an elusive goal, however. Several attempts using magnetic metal contacts to Si[6] or InAs-based two-dimensional electron gas structures[7,8] have resulted in very similar and modest spin-injection effects measured at or below the 1% level. Substantially larger injection efficiencies and related effects are clearly desirable to successfully implement new device concepts. Spin scattering at the interface between the magnetic contact and semiconductor host

*National Research Council Postdoctoral Associate at NRL

appears to be the limiting factor, but very little is known about such interfacial contributions. The small measured effects reported to date also make it difficult to unambiguously identify the factors that affect spin injection.

THE spin-LED

Principle of Operation

We report here the robust and efficient *electrical injection* of spin-polarized carriers from a magnetic contact into a semiconductor heterostructure. This is demonstrated by the fabrication of a spin-polarized light-emitting diode (spin-LED) structure. In a normal LED, electrons and holes recombine in the vicinity of a *p-n* junction or quantum well to produce light when a current flows under forward bias conditions. This light is unpolarized, because all carrier spin states are equally populated, permitting all dipole-allowed radiative transitions to occur with equal probability. In the work reported here, an epilayer of the II-VI semimagnetic semiconductor $Zn_{1-x}Mn_xSe$ is used as the spin-injecting contact to a III-V semiconductor-based LED structure, which consists of a GaAs quantum well flanked by $Al_yGa_{1-y}As$ barrier layers. Figure 1 shows a cross-section of the layered structure. Spin-polarized electrons are electrically injected across the II-VI/III-V heterointerface and into the GaAs quantum well, where they radiatively recombine. Radiative recombination of spin-polarized carriers results in the emission of right (σ-) or left (σ+) circularly polarized light.[3] Polarization analysis of the electroluminescence from such a structure effectively interrogates the spin population of the GaAs quantum well. Thus the existence of circularly polarized electroluminescence demonstrates successful electrical spin injection. A simple analysis of the circular polarization using these selection rules provides a quantitative assessment of injection efficiency. The spin polarization of these electrons deduced from the optical polarization is at least 50%, demonstrating highly efficient spin injection across a heterointerface.

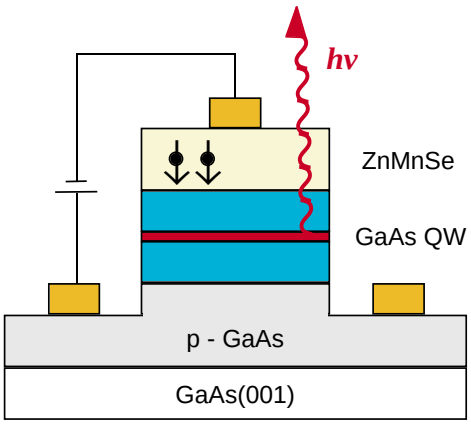


FIGURE 1
Schematic cross section of the samples illustrating the layered structure, contact metallization and voltage biasing during measurement. The layer thicknesses are not to scale.

Semimagnetic semiconductors such as $Zn_{1-x}Mn_xSe$ are well studied, and are noted for the very large band edge spin splitting they exhibit in an applied magnetic field (giant effective Zeeman effect).[9] For modest fields, the spin splitting significantly exceeds *k*T at low temperature. In particular, the splitting of the spin-up ($m_j = +1/2$) and spin-down ($m_j = -1/2$) electron states in $Zn_{0.94}Mn_{0.06}Se$ is approximately 10 meV at 3 Tesla, so that the conduction band effectively forms a completely polarized source of spin-down electrons. This same effect has been used in the past to create a spin superlattice, in which carriers of opposite spin occupy alternating layers of a multilayer structure.[10] Under appropriate electrical bias, these carriers are injected across the ZnMnSe/AlGaAs heterointerface and into the GaAs quantum well. Here, they radiatively recombine with an unpolarized hole population provided by *p*-type doping and emit circularly polarized light.

Device Fabrication

The samples studied were grown on semi-insulating GaAs(001) substrates by molecular beam epitaxy (MBE) in NRL's Epi-Center Facility (Code 6000), a multichamber system for MBE growth, surface analysis, and processing. The growth sequence (Fig. 1) consisted of a 1 µm *p*-type GaAs buffer layer, a 2000 Å *p*-doped AlGaAs barrier, an undoped 150 Å GaAs quantum well, and an *n*(Si)-doped 500 Å AlGaAs barrier. Dopant setbacks of 250 Å were used on either side of the well, and $p(\text{Be}) = 10^{18}$ cm^{-3}. The sample was then transferred to a second MBE chamber connected to the first, where a 2000 Å epilayer of *n*(Cl)-doped $Zn_{0.94}Mn_{0.06}Se$ was grown. This growth was initiated by exposing the (2×4)-As reconstructed surface of the AlGaAs to the Zn flux for 60 s at the growth temperature of 300°C to minimize the formation of defects near the interface.[11] For these growth conditions, the ZnSe/GaAs conduction band (CB) offset is 100 meV,[12] with the ZnSe band edge at higher energy. The band gap of $Zn_{0.94}Mn_{0.06}Se$ is nearly equal to that of ZnSe (2.8 eV at 4.2K) due to bandgap bowing, while that of AlGaAs increases with Al concentration. An Al concentration of 0.1 was chosen for the barrier to minimize the ZnMnSe/

AlGaAs CB offset, which is calculated to be less than 10 meV. The depth of the GaAs CB quantum well is ~100 meV. A doping level of $n = 10^{17}$ cm^{-3} was used for both the AlGaAs and the ZnMnSe. Figure 2 shows a simplified flat band diagram.

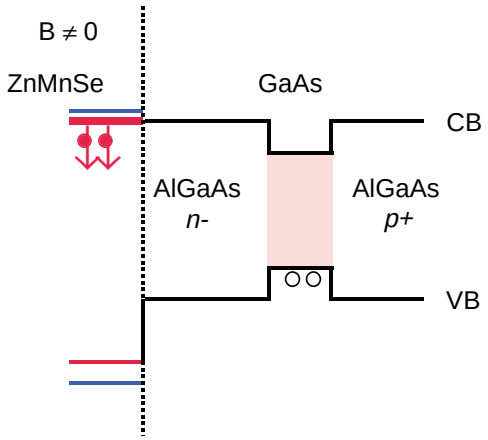


FIGURE 2
Flat band diagram illustrating the conduction and valence band (CB, VB) offsets in the spin-LED structure. The CB of the $Zn_{0.94}Mn_{0.06}Se$ splits with applied field, so that the spin down states are occupied. The holes in the GaAs quantum well are unpolarized. The band offsets are exaggerated for clarity.

The samples were processed into surface-emitting LED mesa structures 200-400 µm in diameter using standard photolithographic techniques. Mesa isolation was achieved by chemically etching to the p-GaAs buffer, and electrical contacts were made to the ZnMnSe and p-GaAs base via Ti/Au liftoff. The top metal contact consists of concentric rings, leaving most of the mesa surface optically transparent. Figure 3 shows the final processed structures.

Demonstration of Electrical Spin Injection

Electrical leads were attached, and the samples were placed in a magnetic cryostat with optical access along the field direction (Faraday geometry). The LED structures were electrically biased to inject electrons from the ZnMnSe into the GaAs quantum well at current densities of ~ 0.01 A/cm^2. The surface-emitted electroluminescence was measured with a spectrometer using a combination of a quarter wave plate and linear polarizer for polarization analysis.

Figure 4 shows representative spectra of the light emitted from the spin-injected LED for selected values of the applied field. The spectra are normalized and aligned to the zero field spectrum to facilitate comparison. At zero field, no optical polarization is observed, as expected, since $Zn_{0.94}Mn_{0.06}Se$ is a Brillouin paramagnet, which acquires a net magnetization only in a magnetic field. The emission peaks near 1.524 eV, which is attributed to recombination with heavy holes. This confirms that radiative recombination occurs in the GaAs quantum well. The polarization rapidly increases with applied field as the spin-polarized electron population is created in the ZnMnSe and injected into the LED structure. The corresponding spectra reveal a significant difference in intensity between the σ+ and σ– components of the electroluminescence, even at 0.5T. Other effects that might contribute to the optical polarization were carefully considered and were either eliminated or included in the error bars. For example, the Faraday rotation resulting from transmission through the ZnMnSe is negligible due to the very short path length (2000 Å), and because the GaAs emission wavelength is far from that corresponding to the band gap of $Zn_{0.94}Mn_{0.06}Se$. Photoluminescence data from the GaAs well show little polarization.

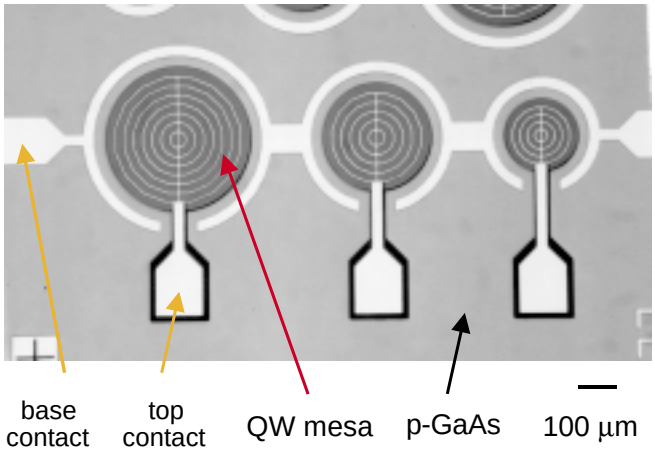


FIGURE 3
The processed surface-emitting LEDs. The top contact bond pad is insulated from the p-GaAs by a layer of SiN.

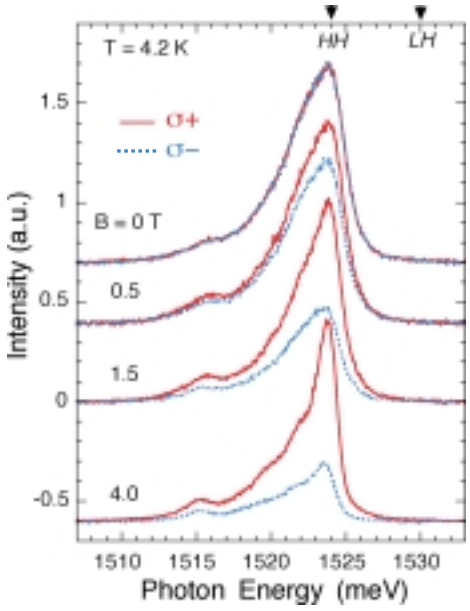


FIGURE 4
Electroluminescence spectra from one of the spin-LED samples for selected values of applied magnetic field, analyzed for left (σ+) and right (σ-) circular polarization. The peaks are normalized and shifted to lower energies by the following amounts to align with the zero field spectrum: 0.5 T (0.11 meV), 1.5T (0.61 meV), 4T (2.15 meV). The relative positions of the heavy and light hole features (*HH*, *LH*) obtained from photoreflectivity (B = 0) are indicated.

At B = 4T, two distinct features are visible at ~ 2 meV and 4 meV to the low energy side of the heavy hole peak. These are consistent with contributions from donor- and acceptor-bound excitons, respectively. These features are more distinct at 8T and are discussed elsewhere. The feature near 1.515 eV is attributed to recombination in bulk GaAs and also exhibits a strong polarization, indicating that some portion of the spin-polarized electrons retain their spin until they recombine in the *p*-GaAs buffer layer. It should be noted that the observation of polarization in such a simple dc measurement clearly demonstrates that the electron spin lifetime is much longer than the radiative recombination time.[4,5]

The degree of circular polarization is obtained from the integrated intensity as P_{circ} = [I(σ+) - I(σ-)]/ [I(σ+) + I(σ-)] and is summarized in Fig. 5 as a function of applied field. P_{circ} saturates around 4T at a value of 50% and decreases slightly at the highest fields. This decrease is not well understood at present but is attributed in part to the Zeeman effect in GaAs, whose spin-splitting is much smaller but opposite in sign to that of $Zn_{0.94}Mn_{0.06}Se$. In a Brillouin paramagnet such as $Zn_{1-x}Mn_xSe$, the dependence of the magnetization (and spin splitting) on applied field and temperature is well described by a Brillouin function.[9,13] The solid line in Fig. 5 shows the spin splitting of $Zn_{0.94}Mn_{0.06}Se$ calculated from a standard Brillouin function analysis and scaled by a multiplicative factor to fit the polarization data. The excellent agreement with the field dependence of the circular polarization for B < 5T confirms that the polarization of the electroluminescence results from spin-polarized electrons that are electrically injected from the $Zn_{0.94}Mn_{0.06}Se$ conduction band and into the GaAs quantum well.

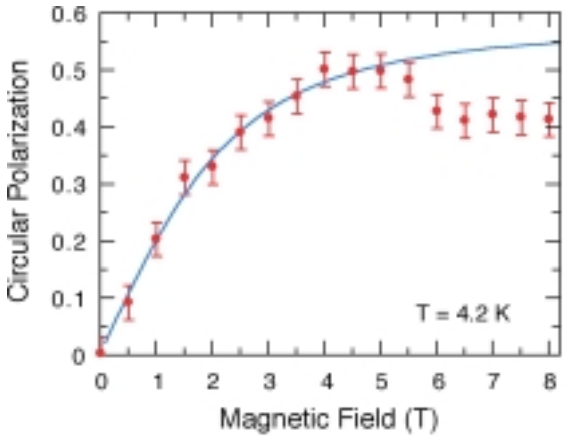


FIGURE 5
Magnetic field dependence of the optical polarization P_{circ}. The solid line is a simple Brillouin function fit to the data.

Quantitative Analysis of Spin Injection Efficiency

Figure 6 illustrates the selection rules that govern the radiative recombination of spin polarized carriers in cubic semiconductors in the Faraday geometry.[3,14] They permit a simple analysis of the data that provides a quantitative measure of the spin polarization of the carriers involved, and hence of spin injection efficiency across the ZnMnSe/AlGaAs interface and into the quantum well. In bulk zincblende semiconductors such as GaAs, the conduction band is two-fold degenerate at the center of the Brillouin zone, corresponding to spin-up and spin-down electrons ($m_j = \pm 1/2$). The valence band is four-fold degenerate, and consists of heavy hole (*HH*) and light hole (*LH*) bands with large and small effective mass, respectively, which are each two-fold spin degenerate ($m_j = \pm 3/2, \pm 1/2$). Radiative electron-hole recombination is allowed for interband transitions that obey the selection rule $\Delta m_j = \pm 1$.[3,14] The probability of a given transition is weighted by the matrix ele-

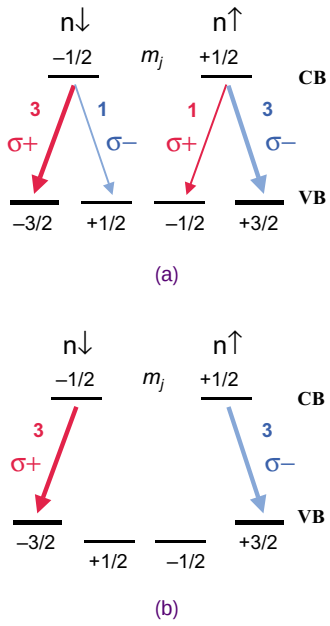


FIGURE 6
Radiative interband transitions allowed by the selection rules for the case of (a) degen-erate, and (b) nondegenerate *HH* and *LH* bands.

ment connecting the levels involved, so that transitions to *HH* states are three times more likely than those to *LH* states.

The net circular polarization of the light emitted can readily be calculated for a given occupation of the quantum well carrier states. Assuming an unpolarized spin-degenerate hole population, a general expression for the degree of circular polarization in the Faraday geometry follows directly from Fig. 6(a) and can be written in terms of the relative populations of the electron spin states n↑ (m_j = + 1/2) and n↓ (m_j = – 1/2), where 0 ≤ n ≤ 1, and n↑ + n↓ = 1:

$$P_{circ} = [I(\sigma+) - I(\sigma-)]/[I(\sigma+) + I(\sigma-)] = 0.5\ (n\downarrow - n\uparrow)/(n\downarrow + n\uparrow). \quad (1)$$

The optical polarization is directly related to the electron spin polarization (n↓ – n↑) / (n↓ + n↑) in the quantum well and has a maximum value of 0.5 due to the degeneracy of the *HH* and *LH* bands. The experimentally measured value P_{circ} = 0.5 suggests spin injection with an efficiency of 100%.

However, the *HH* and *LH* bands are separated in energy by quantum confinement, which modifies Eq. (1) and significantly impacts the analysis. The *HH*/*LH* band splitting is typically several meV, even in shallow quantum wells, and is much larger than the thermal energy at low temperature (~0.36 meV at 4.2 K), so that the *LH* states are at higher energy and are not occupied. For the structures studied here, a calculation that includes corrections for exciton binding energies yields a value of 5 meV for the *HH*/*LH* splitting. This is slightly smaller than the value of 6 meV obtained from photoreflectivity measurements (these positions are indicated in Fig. 4). In this case, only the *HH* levels participate in the radiative recombination process, as shown in Fig. 6(b), and P_{circ} is calculated as before:

$$P_{circ} = (n\downarrow - n\uparrow) / (n\downarrow + n\uparrow). \quad (2)$$

P_{circ} is equal to the electron spin polarization in the well and can be as high as 1.0. Assuming that only spin-down electrons are injected from the ZnMnSe for B > 4T, the experimentally measured value of 0.5 indicates robust electrical spin injection with an efficiency of 50%, i.e., half of the spin-down electrons injected from the ZnMnSe reach the GaAs quantum well without experiencing a scattering event that flips their spin. It should be noted that P_{circ} decreases rapidly for higher current densities, indicating that local heating at the ZnMnSe contact metallization is reducing the initial spin polarization, as expected from the strong temperature dependence of a Brillouin paramagnet. Therefore, the value of 50% should be regarded as a lower bound to the spin injection efficiency.

SUMMARY AND PROSPECTS

In summary, we have demonstrated highly efficient electrical injection of spin-polarized electrons across a II-VI/III-V semiconductor heterointerface and subsequent transport into a quantum well LED structure. Radiative recombination of the carriers results in the emission of circularly polarized light with P_{circ} = 0.5, indicating that at least 50% of the electrons retain their spin state in the process and that the probability of spin-flip scattering is low. An efficient spin-LED[15] effectively couples carrier spin with optical polarization, and may enable the transmission of polarization-encoded information or quantum information processing and transfer. Fabrication of practical devices will benefit from the use of improved contacts and stronger ferromagnets as the spin injector, perhaps including new ferromagnetic semiconductors such as $Ga_{1-x}Mn_xAs$.[16] Future systematic studies of such structures in which the character of the interface or semiconductor layers are modified in a controlled fashion promise to elucidate the mechanisms that contribute to spin scattering and provide a better understanding of spin transport in device structures. This will enable the design of

magnetoelectronic devices[1,2] that use carrier spin as a new degree of freedom to provide new functionality.

ACKNOWLEDGMENTS

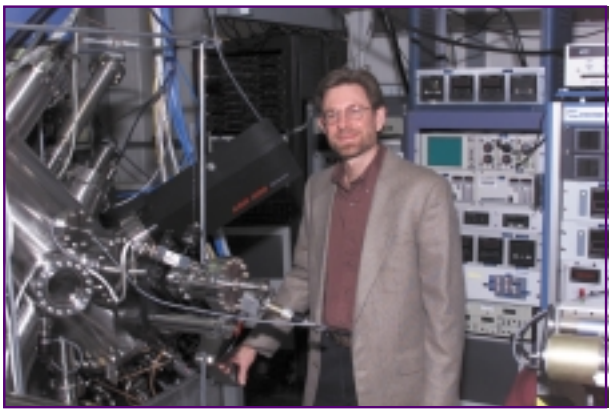
This work was supported by the Office of Naval Research, in part under Research Project 02101, and the DARPA SpinS program. Y.D. Park acknowledges support from ONR as a National Research Council Postdoctoral Associate at the Naval Research Laboratory. The authors gratefully acknowledge helpful discussions with B.V. Shanabrook.

[Sponsored by ONR]

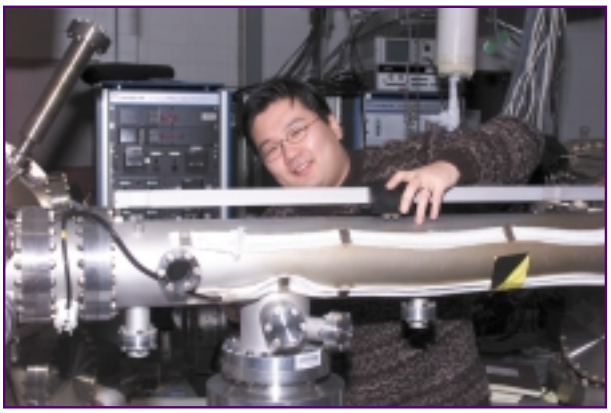
REFERENCES

[1] G.A. Prinz, "Magnetoelectronics," *Science* **282**, 1660 (1998).
[2] S. Datta and B. Das, "The Electronic Analog of the Electro Optic Modulator," *Appl. Phys. Lett.* **56**, 665 (1990).
[3] F. Meier and B.P. Zachachrenya, *Optical Orientation, Modern Problems in Condensed Matter Science* (North-Holland, Amsterdam, 1984), Vol. 8.
[4] D. Hagele, M. Oestreich, W.W. Ruhle, N. Nestle, and K. Eberl, "Spin Transport in GaAs," *Appl. Phys. Lett.* **73**, 1580 (1998).
[5] J.M. Kikkawa and D.D. Awschalom, "Resonant Spin Amplification in n-type GaAs," *Phys. Rev. Lett.* **80**, 4313 (1998); *Nature* **397**, 139 (1999).
[6] Y.Q. Jia, R.C. Shi, and S.Y. Chou, "Spin-Valve Effects in Nickel Silicon/Nickel Junctions," *IEEE Trans. Magn.* **32**, 4707 (1996).
[7] P. Hammer, B.R. Bennett, M.J. Yang, and M. Johnson, "Observation of Spin Injection at a Ferromagnet-semiconductor Interface," *Phys. Rev. Lett.* **83**, 203 (1999).
[8] S. Gardelis, C.G. Smith, C.H.W. Barnes, E.H. Linfield, and D.A. Ritchie, "Spin-valve Effects in a Semiconductor Field-effect Transistor: A Spintronic Device," *Phys. Rev.* **B 60**, 7764 (1999).
[9] J.K. Furdyna and J. Kossut, *Diluted Magnetic Semiconductors*, Vol. 25 of *Semiconductors and Semimetals* (Academic Press, New York, 1988).
[10] W.C. Chou, A. Petrou, J. Warnock, and B.T. Jonker, "Spin Superlattice Behavior in ZnSe / $Zn_{0.99}Fe_{0.01}Se$ Quantum Wells," *Phys. Rev. Lett. 67*, 3820 (1991).
[11] L. H. Kuo, L. Salamanca-Riba, B. J. Wu, G. Hoffler, J. M. DePuydt, and H. Cheng, "Dependence of the Density and Type of Stacking Faults on the Surface Treatment of the Substrate and Growth Mode in ZnS_xSe_{1-x}/ZnSe Buffer Layer/GaAs Heterostructures," *Appl. Phys. Lett.* **67**, 3298 (1995); *J. Vac. Sci. Technol.* **13**, 1694 (1995).
[12] R. Nicolini, L. Vanzetti, G. Mula, G. Bratina, L. Sorba, A. Franciosi, M. Peressi, S. Baroni, R. Resta, A. Baldereschi, J.E. Angelo, and W.W. Gerberich, "Local Interface Composition and Band Discontinuities in Heterovalent Heterostructures," *Phys. Rev. Lett.* **72**, 294 (1994).
[13] J.A. Gaj, "Magnetooptical Properties of Large-gap Diluted Magnetic Semiconductors," in Ref. 9, p. 286.
[14] C. Weisbuch and B. Vinter, *Quantum Semiconductor Structures* (Academic Press, New York, 1991), Chap. 11. [These selection rules may not be rigorously obeyed in real systems due to imperfections or distortion, but provide an analysis good to first-order.]
[15] B.T. Jonker, "Polarized Optical Emission Due to Decay or Recombination of Spin-polarized Injected Carriers," U.S. patent 5,874,749 (February 23, 1999, assignee: U.S. Navy).
[16] H. Ohno, "Making Nonmagnetic Semiconductors Ferromagnetic," *Science* **281**, 951 (1998).

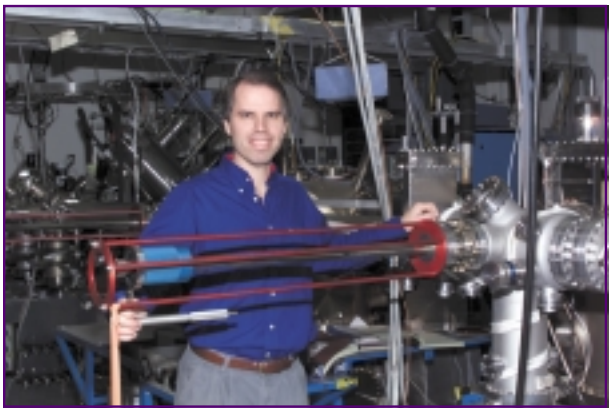
THE AUTHORS



BEREND T. JONKER received a B.A. (Honors) degree in physics from Calvin College in 1977, and an M.S. and Ph.D. degrees in physics from the University of Maryland in 1981 and 1983, respectively. He came to NRL as a National Research Council Postdoctoral Fellow in 1984 and became an NRL staff member in 1986. His research has focused on spin-dependent carrier localization in semiconductor superlattices, electrical spin injection and transport in semiconductor heterostructures, and the fabrication of prototype magnetoelectronic devices. His most recent work addresses the MBE growth and study of III-V magnetic semiconductors. Dr. Jonker has authored 135 scientific publications, holds three U.S. patents with two additional pending, and is a Fellow of the American Vacuum Society.



YUN D. PARK graduated with a B.S. in electrical engineering from Cornell University in 1994, and received his M.S. (1997) and Ph.D. (1998) in materials science and engineering from the University of Florida. Dr. Park's thesis explored the effects of processing and dimensionality of nanometer-sized magnetic structures, from which he authored or co-authored more than twenty scientific journal articles. Soon after graduation, Dr. Park was awarded a National Research Council Post-doctoral Fellowship at the Naval Research Laboratory, where he is currently working with Dr. B.T. Jonker in the Materials Physics Branch in the areas of magnetoelectronic devices, molecular beam epitaxy, thin-film growth, and characterization of magnetic semiconductors.



BRIAN R. BENNETT received the B.S. and M.S. degrees in 1984 and 1985, respectively, from the Massachusetts Institute of Technology. In 1992, he received the Ph.D. degree in materials science and engineering from M.I.T. Since 1992, Dr. Bennett has been at the Naval Research Laboratory. His research focuses on the growth and applications of arsenide, antimonide, and nitride semiconductor heterostructures, including studies of interfacial control, self-assembled quantum dots, field-effect transistors, resonant tunneling diodes, and infrared detectors. Dr. Bennett has published more than 100 technical papers in refereed journals. He also serves as an advisor to postdoctoral associates at NRL.

HAI-DU CHEONG received B.A. and M.S. degrees in physics from Kyungpook National University at Taegu, Korea in 1986 and 1988, respectively, and a Ph.D. in physics from State University of New York at Buffalo in 2000. He is presently a post-doctoral researcher in the Department of Materials Science and Engineering of Pohang University of Science and Technology, Korea. His research is focused on ZnO thin film growth using Laser-MBE and MOCVD techniques, optical characterization of III-V (GaAs-based) and II-VI (Zn-based) semiconductors, and time-resolved photoluminescence spectroscopy. Dr. Cheong has authored six scientific publications.

GEORGE KIOSEOGLOU received B.S and M.S degrees in physics from the Aristoteleio University, Thessaloniki, Greece, in 1987 and 1991, respectively, and a Ph.D from the State University of New York (SUNY) at Buffalo in 1999. He is currently a postdoctoral associate at SUNY Buffalo. His research involves the study of the magneto-optical properties of semiconductors and heterostructures based on them. His recent work is on electrical spin injection in GaAs-based LEDs and X-ray studies of interface quality and intermixing in semiconductor heterostructures. Dr. Kioseoglou is co-author on more than 20 publications in scientific journals.

ATHOS PETROU received a B.S. degree in physics from the University of Athens, Greece, in 1976 and M.S. and Ph.D. degrees from Purdue University in 1979 and 1983, respectively. Between 1983 and 1985 he was a research associate at Northeastern University. He joined the physics faculty at the State University of New York Buffalo in 1985. His research involves the study of the magneto-optical properties of semiconductors and heterostructures based on them. His most recent work is on electrical spin injection in GaAs-based LEDs. Dr. Petrou is a co-author of 87 publications in scientific journals and is a Fellow of the American Physical Society.

Numerical Simulations of Pulsed Detonation Engines

K. Kailasanath, C. Li, and G. Patnaik
Laboratory for Computational Physics and Fluid Dynamics

Detonations are an extremely efficient means of burning a fuel-air mixture and converting its chemical energy content into mechanical energy. Air-breathing and rocket engines based on pulsed detonations have the potential to provide the Navy with increased range and speed while reducing fuel consumption and system costs. Over the past two years, we have conducted extensive computational studies of the pulsed detonation engine concept. The results of these numerical simulations have been invaluable in developing a basic understanding of these engines. New methods for enhancing the performance of these engines have also been identified using the simulations.

INTRODUCTION

In the quest for propulsion systems with reduced acquisition costs, better efficiency, increased range, and reduced fuel consumption, pulsed detonation engines (PDEs) are a potentially revolutionary concept. The mechanical simplicity of the system will translate into reduced production and maintenance costs. Detonations are an extremely efficient means of converting a fuel into products and releasing the fuel's chemical energy content. This intrinsic advantage of detonations results in better thermodynamic efficiency. Therefore, air-breathing and rocket engines based on pulsed detonations have the potential to provide the Navy and DoD with increased range and speed while reducing fuel consumption and system costs. However, detonations have been explored for propulsion only recently.[1] The problems involved in rapidly mixing the fuel and air at high flight speeds and initiating and sustaining a detonation in a controlled manner in fuel-air mixtures have made this difficult. By focusing on pulsed systems rather than continuous systems, additional time is available between the pulses for fuel-air mixing. Several basic research issues need to be addressed in order to mature the PDE technology base. Recent advances in combustion diagnostics and computational combustion will aid in the development of a better understanding of PDEs. This article is an overview of our computational studies of pulsed detonation engines.

WHY DETONATIONS?

Very rapid material and energy conversion is a key feature of detonations. This rapid "burning" or material conversion rate, typically tens of thousands of times faster than in a flame, can lead to several advantages for propulsion, such as more compact and efficient systems. Because of the rapidity of the process, there is not enough time for pressure equilibration, and the overall process is thermodynamically closer to a constant volume process (like an explosion in a closed container) than the constant pressure process typical of conventional jet propulsion systems. To illustrate this point, three idealized thermodynamic cycles are compared in Fig. 1.

For comparison, the only process that is different in the three cycles is the mode of energy conversion or heat addition. For the three cases, heat is added at constant pressure, constant volume, or in a detonation. Hence, the three cycles have been referred to as "constant pressure," "constant volume," and "detonation" cycle, respectively. The amount of heat added is kept the same for the three cycles. In all cases, the fuel-air mixture is initially compressed "adiabatically" (without any heat loss) from 1 to 3 atm. before heat addition. After heat addition, the products of combustion are expanded adiabatically to 1 atm. Finally the system is returned to its initial state. The thermodynamic efficiency is calculated as the ratio of work output to the heat input. Since the

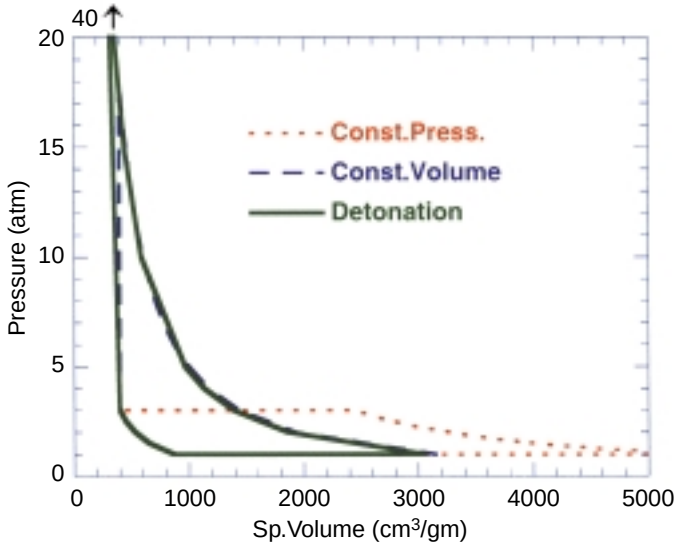


FIGURE 1
Comparison of thermodynamic cycles for constant pressure, constant volume, and detonation modes of combustion.

heat addition was maintained the same for the three processes, the differences in the work done give the relative thermodynamic efficiency of the three combustion processes. The thermodynamic efficiencies for the three cycles are calculated to be 27% for constant pressure, 47% for constant volume, and 49% for detonation. From the above numbers we see that the thermodynamic efficiency of the detonation cycle is significantly better than that of the constant pressure cycle (usually called the Brayton cycle) that is typical of combustion in various forms of jet propulsion. The detonation cycle is closer to a constant volume cycle (usually called a Humphrey cycle) except for a reduction in specific volume and increase in peak pressures. A major challenge in the development of the pulsed detonation engine is attaining this higher potential efficiency in a practical device.

PROCESSES IN A PULSED DETONATION ENGINE

The basic concept behind a pulsed detonation engine is fairly simple and is illustrated in Fig. 2 using a pipe or a tube closed at one end and open at the other. The various colors in the figure correspond to different levels of pressure, with red being the highest (30 atm.) and the violet (or purple) being the lowest (1 atm.). The tube is initially filled with a fuel-air mixture at 1 atm. A shock wave is initiated at the closed end. Chemical reactions in the shock-heated mixture generate pressure waves that couple with the shock wave to form a detonation. After a brief transition period, the detonation travels toward the open end of the tube at a nearly constant velocity called the Chapman-Jouguet (CJ) velocity. This CJ velocity and the corresponding pressure behind the detonation wave (CJ pressure) are characteristics of the fuel-air mixture. The detonation initiation process also generates expansion waves that travel toward the closed end (called head-end) of the tube. As the detonation leaves the tube, additional expansion waves form at the open end and travel toward the closed end. These expansion waves are important because they help evacuate the tube. Once the tube is evacuated, fresh fuel and air are injected into the tube. When the injection and mixing is completed, a detonation is initiated again at the closed end and the whole process is repeated. The cyclical nature of the entire process is what gives rise to the term "pulsed detonation engine (PDE)."

The propulsive thrust and other measures of the performance of such an idealized "engine" can be determined either from the momentum of the gases flowing out of the tube or the pressure at the closed end of the tube. The pressure distributions used in Fig. 2 for illustration purposes were actually values obtained from numerical simulations of an engine operating on a hydrogen-air mixture. Let us briefly look at the computational tools needed to conduct such studies.

COMPUTATIONAL TOOLS

Computational tools are being developed to analyze various problems related to PDEs such as fuel-air mixing, detonation initiation and transmission, and multiphase flow effects. The geometric complexity of the system being simulated as well as the details included in the chemical and physical models vary from problem to problem. For example, in our three-dimensional simulations of fuel-air mixing and initiation of detonations, either no chemistry model or

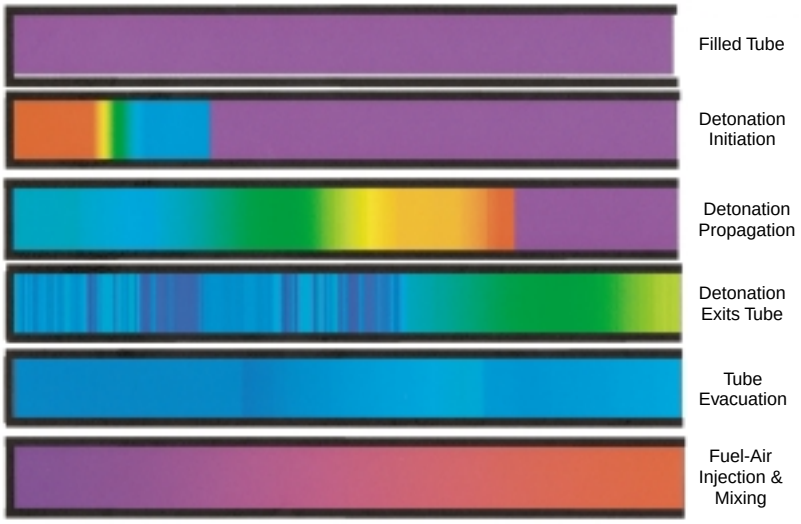


FIGURE 2
Processes in a pulsed detonation engine cycle.

simple models are included, but the details of the geometrical complexity representative of experimental configurations are included. At the other extreme, detailed multi-step chemistry and thermodynamics is included in simulations of geometrically simple systems. We also plan to include simplified chemical kinetic mechanisms of hydrocarbon combustion in some of these simulations. Next we briefly discuss the model and solution approach used in our basic study of the gas-phase reactive flow in a simple PDE.

The time-dependent, compressible, reactive flow equations for the conservation of mass, momentum and energy are solved. The terms in the conservation equations representing the different physical and chemical processes are solved separately and coupled using timestep splitting techniques. This procedure allows the individual processes to be integrated by appropriate and efficient techniques and also permits the easy substitution and elimination of different submodels for the individual processes, as needed for specific applications.

The code can compute the three-dimensional multi-species reactive flow in a simple geometry and allows for the inclusion of species diffusion, thermal conduction, and radiative heat loss. For the short duration, single-cycle simulations discussed here, the diffusive and thermal processes will have a negligible effect and hence only the convective flow with detailed chemistry is considered. For simulations such as those used in Fig. 2, the code is used to compute the one-dimensional flow, neglecting the variations in the other two directions.

For the high-speed reactive flows that are typical of detonations, an explicit algorithm such as the Flux-Corrected Transport (FCT) is very efficient and accurate for integrating the fluid convection. FCT is a conservative, monotonic algorithm with fourth-order phase accuracy. With various initial and boundary conditions, this algorithm has been used previously to solve a wide variety of problems involving detonations.

A comprehensive model for hydrogen combustion with 8 species and 48 reactions or a simplified two-step model for ethylene combustion is used for the results discussed here. Because of the complexity of the reaction scheme and the large number of computational cells in a multidimensional calculation, the solution of the chemical rate equations can take a large fraction of the total computational time. Hence, the code has been developed directly for massively-parallel computers.

PERFORMANCE OF A PDE

As discussed before, a fundamental reason for investigating PDEs is the higher thermodynamic efficiency of a detonation cycle when compared to the constant pressure and the constant volume cycles. The higher efficiency is primarily due to the higher pressures and temperatures attained during the detonation cycle. Pressure profiles and histories at various locations within a PDE have been analyzed to better understand the various factors that may control the performance. The results from one-dimensional numerical simulations have been used to generate a "typical" profile of the head-end (closed-end) pressure shown in Fig. 3. Measures of the propulsive performance such as thrust can be calculated from such information. In Fig.3, the abscissa is normalized using the detonation transit time (tube length/CJ detonation velocity) and the ordinate is normalized using the CJ detonation pressure.

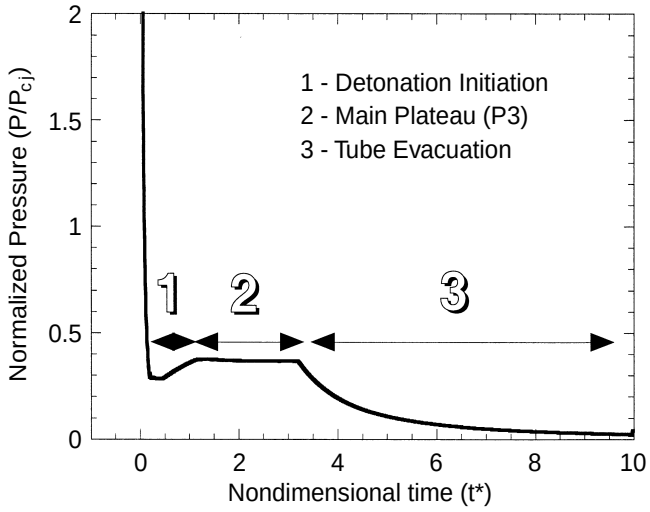


FIGURE 3
Different stages in the time evolution of the pressure at the closed end of a PDE tube.

In Fig. 3, we have divided the time history into three distinct stages or regimes: initiation, plateau, and relaxation. Different factors control the different stages. The first stage is dominated by the method used to initiate the detonations and to some extent the details of the transition process. Attention is usually focused on the second, the plateau region or stage, because the value of the plateau pressure can be related to the CJ detonation pressure and hence is a property of the fuel-air mixture. The third stage describes the relaxation of the plateau pressure to the ambient value. Until recently, this stage has not been studied in detail.

ROLE OF PRESSURE RELAXATION

We have conducted numerical simulations to investigate the role of the rate of pressure relaxation to the ambient conditions. Numerically, for the idealized straight tube PDE, this can be varied by the specification of the boundary conditions at the open end of the tube. Several choices for the open boundary condition have been tried and their effect on the flow field and performance evaluated. For the cases discussed here, a boundary condition implementation based on the method of characteristics is used. This ensures that no constraints are imposed on the flow quantities when the outflow is supersonic and enforces the required constraints when the flow becomes subsonic. Even in this formulation, there remains a free parameter in the subsonic case that needs to be specified. Various choices for this parameter result in different rates of relaxation for the pressure at the open boundary.

The results for three choices of the relaxation parameter (slow, intermediate and fast) are shown in Fig. 4. Clearly, the area enclosed by the curves, which is a measure of the thrust of the engine, is different for the three cases. The time integral of the thrust gives the impulse. The impulse per unit mass defines a quantity called "specific impulse" (I_{sp}), which is usually taken as a standard measure of the propulsive performance of an engine. It has the same unit, seconds, in the various systems of units. The numerical simulations show that the peak I_{sp} is 60% larger for a slow relaxation process than for a fast one. In practice, different relaxation rates may be attained by suitably tailoring the shape of the exhaust nozzle.

EFFECTS OF BYPASS AIR

In another series of simulations, a 50-cm long tube was filled with an ethylene-air mixture to various lengths and the rest of the tube was filled with air. Detonations were initiated at the closed end, as described earlier. Figure 5 shows the time history of the impulse for various cases. An interesting observation is that the impulse is not proportional to the amount of fuel fill. When the degree of fill is decreased from 100% to 20%, the peak impulse decreases from 604 to 381 N-s/m^2. That is a decrease of only 37%. Detailed analysis of these multidimensional simulations[2] shows that this result is due to the presence of two different sets of expansion waves, one from the fuel-air interface and the other from the exit-end of the tube. When these different sets of expansion waves reach the thrust wall, the pressure decays at

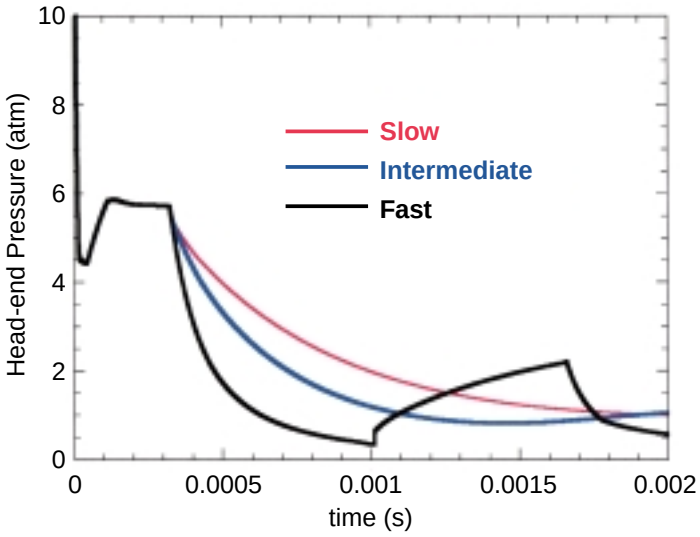


FIGURE 4
Effect of different pressure relaxation conditions at the open end on the pressure evolution at the closed end of a PDE tube.

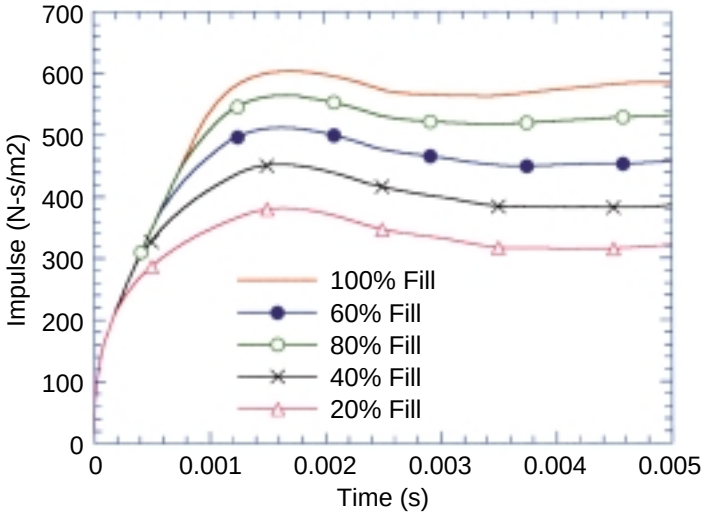


FIGURE 5
Effects of partial filling of the PDE tube with fuel-air mixture on the impulse generated.

different rates because the strength of these expansion waves is different.

This result has several implications. It provides a means of controlling the thrust by controlling the amount of fuel-air fill in the detonation chamber. Furthermore, it suggests that a significant performance drop may not occur if, during multi-cycle operations, the tube is not filled completely. There is also the intriguing possibility of enhancing the performance by using bypass air (the air that would otherwise go around the engine) during flight. This is similar to an idea that is currently used in conventional jet engines, where a major improvement in performance is achieved by effectively using some of the air that would normally bypass (flow around) the engine.

DETONATION DIFFRACTION SIMULATIONS

The large amount of energy required to directly initiate detonations in large tubes or chambers is considered to be one of the potential limitations of the PDE concept. Several techniques have been looked at to circumvent this problem. In one approach, detonation is initiated in a small tube (which requires less

energy) and then allowed to diffract (travel through an area change) into the larger tube or chamber. Preliminary experimental studies at the Naval Postgraduate School (NPS) in Monterey, California, showed that even with this technique, it was difficult to initiate a detonation if both tubes were filled with the same fuel-air mixture such as ethylene and air. However, if the smaller tube had an ethylene-oxygen mixture, it was possible to initiate detonations in the larger tube, even when it contained an ethylene-air mixture. We conducted multidimensional numerical simulations to understand and explain these observations.[3]

The front of any propagating detonation is not a straight line but has a complex shape (referred to as "detonation structure") with "triple points," where three different shock waves meet. The detonation structure in the smaller tube as well as the structures in the larger tube at a series of times are shown in Fig.6 using pressure distribution plots. As seen in the figure, when both tubes are filled with the same ethylene-oxygen mixture, a detonation successfully transmits from the smaller tube into the larger tube. There are eight pairs of triple points in the smaller tube shown in Fig. 6. As the detonation diffracts into the larger tube, expansion waves from the area change weaken the detonation structure causing some decoupling between the reaction fronts and the leading shock waves. Some of the triple points are seen to disappear by 130 microseconds after initiation in the smaller channel. However, new triple points are generated in time and the detonation is seen to recover from the effects of the area change by 165 microseconds. Note that the detonation wave has not yet reached the walls at this time. Some changes in the front are observed as the detonation front reflects from the walls but by now, a self-sustained propagating detonation has been created in the larger tube.

The picture is quite different when both tubes are filled with the same ethylene-air mixture. In this case, the detonation does not survive the diffraction into the larger tube, as seen in Fig. 7, where the pressure distribution in the larger tube is shown at a particular time (200 microseconds after initiation in the smaller channel). The front is devoid of the wave structures that are characteristic of propagating detonations.

The marked difference in the two cases can be traced to the differences in the number of triple points in the two mixtures. Our simulations of detonation structures in these ethylene-oxygen and ethylene-air mixtures showed that a key difference between the two detonations was the wider spacing between triple points for the air diluted mixtures. That is, a tube of a given size can have only fewer triple points in the air mixture than in the oxygen mixture. When a detonation in an ethylene-air mixture in the small tube (with few triple points) travels into the larger tube, it does not appear to be able to retain enough triple points to initiate a propagating detonation in the larger chamber. We can increase the number of triple points in the smaller channel by filling it with an oxygen-enriched mixture. Indeed, if we use the ethylene-oxygen mixture in the smaller tube, the detonation successfully makes the transition into the ethylene-air mixture in the larger tube as shown in Fig. 8.

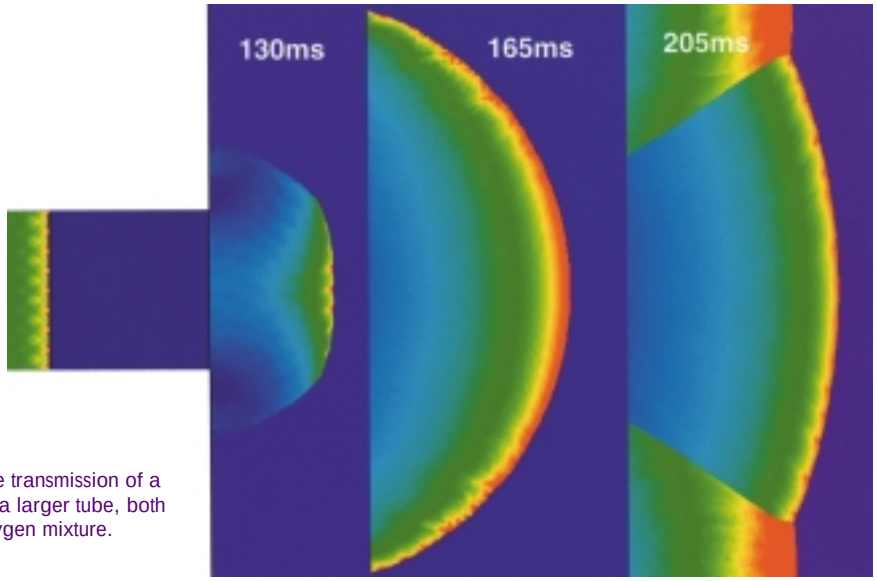


FIGURE 6
Pressure distribution showing the transmission of a detonation from a small tube to a larger tube, both filled with the same ethylene-oxygen mixture.

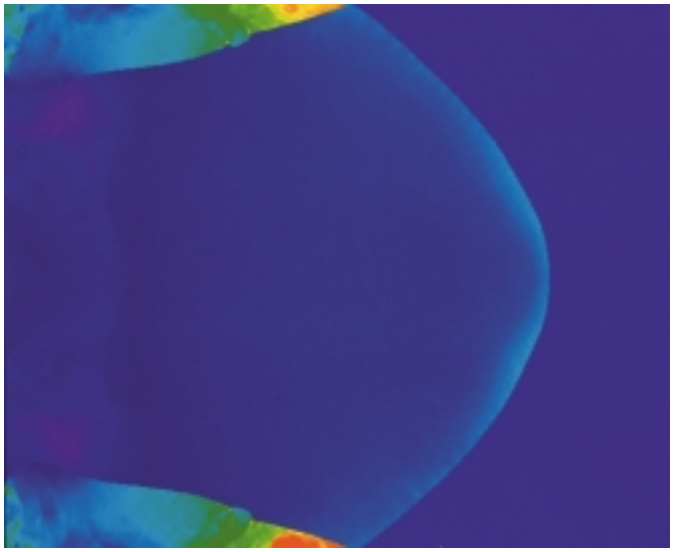


FIGURE 7
Pressure distribution in a large tube after unsuccessful transmission of a detonation from a smaller tube. Both tubes were filled with the same ethylene-air mixture.

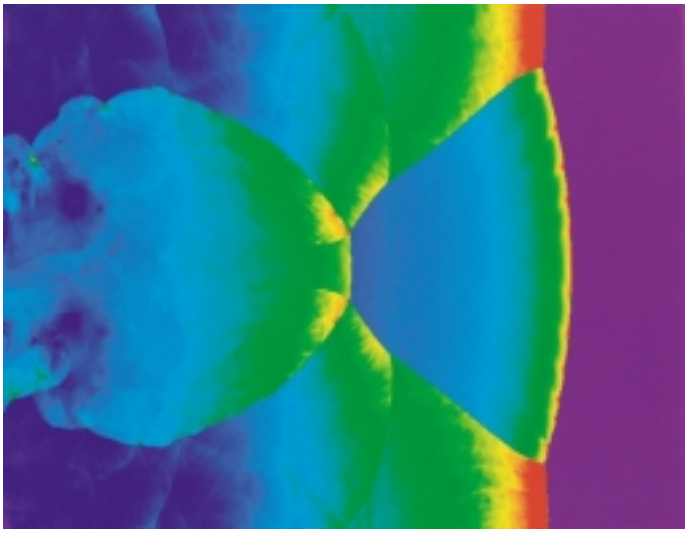


FIGURE 8
Pressure distribution showing the detonation structure in a large tube filled with an ethylene-air mixture. The detonation was initiated using a smaller tube filled with an ethylene-oxygen mixture

SUMMARY

In this article, the impact of numerical simulations on the research of the pulsed detonation engine concept has been highlighted using several examples. A fundamental understanding of various factors that control the performance of PDEs has been attained from the computational studies. This basic understanding has been used to explore means of enhancing the performance. Partially filling the tube with air is found to be one such option. The performance can also be increased by slowing the rate of pressure relaxation at the exhaust nozzle of the engine. The role of initiators is explored using multidimensional simulations. The experimental difficulties found in initiating detonations in ethylene-air mixtures has been traced to the low number of triple points in "typical" initiator tubes. The ease of initiation with fuel-oxygen mixtures has been demonstrated. The major emphasis of our current and future work will be on liquid-fueled pulsed detonation engines, because it is more advantageous to use liquid fuels in many volume-limited Naval applications. Specific issues that are being considered include the effects of droplet size, atomization, vaporization and multiphase mixing on the detonation initiation, propagation, structure and failure.

ACKNOWLEDGMENTS

This work is sponsored by the Mechanics and Energy Conversion S&T Division of the Office of Naval Research.

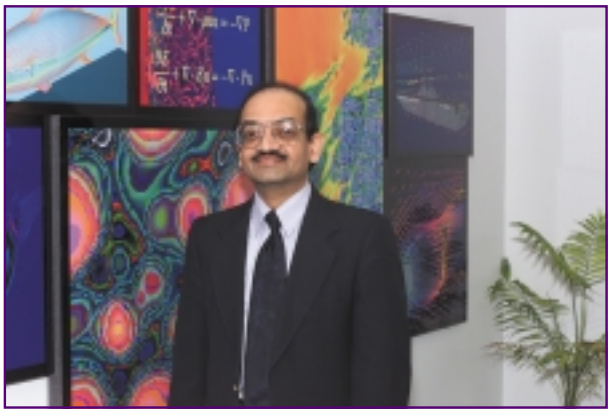
REFERENCES

[1] Kailasanath, K., "Review of Propulsion Applications of Detonation Waves," AIAA Journal, Vol. 38, No.9, pp. 1698-1708 (2000).

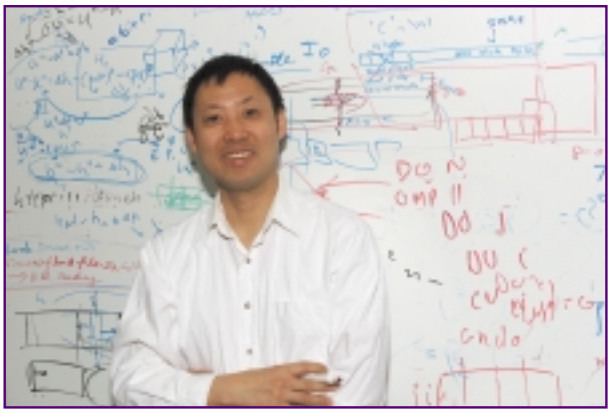
[2] Li, C., Kailasanath, K., and Patnaik, G., "A Numerical Study of Flow Field Evolution in a Pulse Detonation Engine," AIAA Paper 2000-0314, American Institute of Aeronautics and Astronautics, Reston, VA, January 2000.

[3] Li, C., and Kailasanath, K., "Detonation Diffraction in Pulse Detonation Engines," AIAA Paper 2000-3470, American Institute of Aeronautics and Astronautics, Reston, VA, July 2000.

THE AUTHORS



KAZHIKATHRA KAILASANATH is the Head of the Center for Reactive Flow and Dynamical Systems in the Laboratory for Computational Physics and Fluid Dynamics. He received his B.Tech degree (1976) from the Indian Institute of Technology, Madras, and M.S.A.E. (1979), and PH.D. (1980) degrees from the Georgia Institute of Technology. He has been at NRL since 1980 and became Head of the Center in 1988. His research interests include the structure, stability and dynamics of flames and detonations; combustion instabilities, multiphase flows; and the simulation of advanced propulsion system concepts. He is a Fellow of the Institute of Physics, an Associate Fellow of the AIAA, and a member of The Combustion Institute and the American Physical Society.



CHIPING LI graduated from the University of California, San Diego, with a Ph.D. degree in engineering sciences in 1987. He has worked at the Laboratory for Computational Physics and Fluid Dynamics in the Naval Research Laboratory as a research scientist since 1989. His areas of research include numerical simulation of reactive flows, basic combustion processes, and algorithms for parallel computation. His recent research activities focus on numerically studying combustion processes such as in supersonic flows including detonation and other types of shock-induced combustion. Dr. Li's work in this area is closely related to propulsion applications such as the ram accelerator and the pulse detonation engine.

GOPAL PATNAIK received his B.E. (Honors) in mechanical engineering in 1980 from the Birla Institute of Technology and Science, India, and his M.S. (1981) and Ph.D (1986) in mechanical engineering from Carnegie-Mellon University. Since 1982, Dr. Patnaik has been involved in the development and application of state-of-the-art computational techniques for low-speed reactive flows. Since 1997, as a member of the Laboratory for Computational Physics and Fluid Dynamics, he has used parallel computing techniques to study complex three-dimensional flames. Dr. Patnaik has published several papers and made numerous presentations on the subject. He is a member of the Combustion Institute and a senior member of AIAA.

Phase-Coherent Underwater Acoustic Communications: Building a High-Data-Rate Wireless Communication Network in the Ocean

T.C. Yang
Acoustics Division

Oceanographic and underwater acoustic research today is limited by the expense of data telemetry using cables. Naval operations are requiring that increasingly large amounts of data be transferred between submerged platforms, surveillance sensor nodes, and surface ships. This demands high-data-rate underwater acoustic communications that, until recently, have not been possible because of the time space variable nature of the underwater acoustic channel—a fading channel with severe time and frequency spread. A joint adaptive decision feedback equalizer with a phase-locked loop has recently been developed that is capable of adapting to channel variation, thereby enabling high-data-rate phase-shifted-keying communications (ACOMMS) in high signal-to-noise (SNR) propagation conditions. Real-time implementation of these algorithms has been demonstrated using high-speed digital signal processing chips. In support of Office of Naval Research ACOMMS efforts, NRL has coupled random media acoustic propagation physics simulators with ACCOM algorithms to develop realistic acoustic communications performance prediction models and improve the efficiency and adaptability of the ACCOMS processing algorithms. The studies have been supported with at-sea acoustic and environmental experiments. Our efforts to understand the impact of random media acoustic propagation on acoustic communications processing algorithms have resulted in significant breakthroughs. These breakthroughs include the development of new signal processing approaches that reduce bit error rates in low SNR propagation conditions. This article discusses the difficulties of using the underwater acoustic channel, the recent advances in communication technology, and the future of the wireless underwater communications for network-centric warfare and ocean exploration.

INTRODUCTION

The ocean is a frontier where accelerated scientific and commercial exploration can be expected in the 21st century. Acoustical oceanography uses acoustics as a means for ocean study. Weather forecasting requires sensor data, not just from the ocean surface but also from within the ocean water column. Acoustic surveillance of the ocean for hostile submarines requires real-time in situ environmental acoustic data to localize the source and maximize the detection range of Navy systems. Environmental acoustic data collected with today's technology are too sparsely sampled to meet future needs. The bottleneck lies predominantly in the undersea telemetry link; current systems use cables that are expensive and immobile. The future, however, looks promising with the advent of wireless undersea acoustic communications (ACOMMS). The availability of inexpensive acoustic sensors and fast digital processing chips can make an undersea wireless communication network a reality. Using the analogy of traffic reporting around the Beltway, where previously information came primarily from aerial surveillance by a few helicopters, real-time traffic updates today come from drivers at the scene using mobile cellular phones. This makes the use of an expensive helicopter less necessary.

The Navy has long recognized the need for high-data-rate underwater acoustic communications. A critical need is a high-data-rate undersea communication link between submarines and surface ships at a distance. Currently, the Navy uses an underwater telephone whose data rate is limited by the multipath propagation conditions commonly experienced in the ocean. Communications suffer from a long series of echoes to the extent that messages sometimes be-

come unintelligible. Under the sponsorship of the Office of Naval Research (ONR) and the Defense Advanced Research Projects Agency (DARPA), a digital high-data-rate, band-efficient, phase-coherent acoustic communication technology has recently been demonstrated by the Woods Hole Oceanographic Institution and Northeastern University. To achieve high-data rate for quadrature phase-shifted keying (QPSK) signals, the investigators implemented a jointly adaptive decision feedback equalizer (DFE) with a phased-locked loop (PLL) to adapt to the time- and space-varying underwater acoustic channel (see discussions below). Recent sea tests (e.g., ACCOMMS ATD, see below) have demonstrated data transmission at voice rate (2 kbits/s) at long ranges (> 30 km) using a midfrequency (2-5 kHz) band signal and transmission at video rate (20 kbits/s) at close ranges (2-5 km) using a high-frequency (10-30 kHz) band signal. The processing algorithms are computationally intensive. Real-time implementation of these algorithms requires the use of high-speed signal signal processing (DSP) chips. Although some technical problems remain to be solved, the feasibility of real-time high-data-rate undersea communications has opened the doors to new approaches to undersea and mine countermeasure warfare, to oceanographic studies, and to marine industrial applications.

Tactical Oceanography

Currently, information about the ocean is obtained from ships, satellites, floats, and moorings. Sensors such as thermistor chains and acoustic Doppler current profilers, either deployed from ships or moored on the ocean bottom, provide quasi-synoptic data across a two-dimensional (2D) (depth/range or depth/time) section through the evolving fields. Satellites provide a 2D measurement of ocean surface properties. With only limited data from a small number of platforms and satellites, extrapolation of the data to a 3D field or to a higher resolution grid introduces temporal/spatial aliasing. The technology push is toward a high-resolution sampling of the ocean: High-resolution data are required to understand various oceanographic processes. These processes include front dynamics, where cross-front circulation must be resolved, and surface layer dynamics, where the mechanism of mixing must be resolved.

High-resolution sampling of the ocean requires linking many sensors together. The ONR project, Autonomous Oceanographic Sampling Network (AOSN), attempts to localize oceanographic processes by using a tomographic approach and conducts detailed sampling using autonomous undersea vehicles (AUVs). Wireless undersea acoustic communications are used for command and control of AUVs and for data telemetry between the AUVs and other acoustic sensor nodes.

In the future, ocean sampling could be accomplished with a network of inexpensive sensors. Without the need to lay cable, the cost of data transmission would be significantly reduced. Significant number of sensors could be deployed and used to build an underwater network analogous to the computer Internet system.

Undersea Surveillance

In the past, undersea surveillance used permanently deployed systems (such as SOSUS). These systems are expensive and require long lead times for laying and burying cable. For protection of battle groups in hostile shallow-water areas, the emphasis is on readily deployable array systems (RDS) that can last several months. In forward areas, clandestine deployment and covert operation are required to avoid compromising the sensor system and the supported activity. Systems that have a surface expression risk easy detection, and systems that use cables risk damage by incidental trawling and anchoring. Both systems can be defeated by low-technology dragging or grappling. Wireless undersea communication systems eliminate the vulnerable telemetry hardware and avoid conspicuous emplacement.

The establishment of an acoustic local area network using wireless undersea acoustic communications could transform undersea surveillance in the future. Using a local area acoustic network, all undersea sensors could be linked together. Undersea connectivity could be a critical part of the future net-centric warfare system that extends from undersea to above sea. Target detection and localization information from undersea surveillance systems could be transmitted to friendly submarines operating in the area to achieve critical tactical advantages. An undersea network of sensors could be used as a barrier system for protecting harbors and critical choke points. Asynchronous parallel access to the network and the ability to communicate either covertly or with a low probability of interception are two research areas of great interest.

UUVs, Mine Countermeasures, and Tactical Operations

High-data-rate underwater acoustic communications provide the critical undersea data link between unmanned underwater vehicles (UUVs), submarines, and surface ships. The ability to acoustically communicate between a submarine and a surface ship is critical for coordination of tactical operations for battle

group defense. The ability to remotely command and control (downlink) and retrieve (uplink) data from UUVs using a wireless acoustic link will significantly enhance the capability of mine countermeasures (MCM) and covert operations. Transferring such duties from a remotely operated (tethered) vehicle to a UUV eliminates the risk of exposure to the mother ship and allows multiple vehicles to operate. In forward areas, in areas not accessible by conventional platforms, and in denied areas that present undue risks to manned systems, UUVs offer unique capabilities in reconnaissance and intelligence gathering due to their autonomy, low observability, deployability, and environmental adaptability. For command and control of UUVs from surface ships or submarines, a tactical midfrequency band (2-5 kHz) is used that extends horizontal telemetry ranges to 10-30 km and beyond. For high-data-rate communications between UUVs and a mother ship, a high-frequency band (10-30 kHz) will be used over short (2-5 km) ranges. The ability to acoustically communicate between Navy platforms (using phase-coherent technology) has recently been demonstrated by the Tactical Acoustic Communications Advanced Technology Demonstration (ATD) project sponsored by the Submarine Warfare Division in the Office of the Chief of Naval Operations. The use of high-data-rate acoustic communications for UUVs in future MCM missions has received great interest recently from the Navy.

Outline

We first overview the basic challenges of underwater acoustic communications and the recent advances made at NRL. We then address both the design of signal processing algorithms and the impact of environmental variability on the algorithm performance. The NRL approach to the ACOMMS problem is based on the physics of random media propagation and the resultant impact on the complex acoustic signal properties most likely to impact ACOMMS performance. Signal processing algorithms are next developed to mitigate the signal variability. Recent accomplishments and advances that are unique to this approach are described. Future research efforts and future technology development (toward a high-data-rate acoustic local area network) are addressed.

ENVIRONMENTAL INFLUENCES

The ocean poses a unique temporal and spatial variable propagation environment that is very different from the radio-frequency (RF) propagation environment experienced by the cellular phone network. Wireless communications in the ocean must be done acoustically since RF signals travel only tens of meters in seawater. High-frequency sound travels shorter distances than low-frequency sound because of sound absorption by seawater. Consequently, the carrier frequency and the available bandwidth for wireless acoustic communications will be range-dependent. To increase the data rate, band-efficient communications using modulation schemes such as phase-shifted-keying are preferred to the phase-incoherent schemes using frequency-shifted-keying. However, the performance of phase-coherent schemes is often environment-dependent. For less favorable environments (such as the so-called overspread channel), it may be necessary to consider phase-incoherent schemes for robustness.

When will phase-coherent acoustic communications work? We briefly summarize the technical challenges of underwater acoustic communications and discuss the environmental acoustic parameters that impact the algorithm performance.

Underwater Acoustic Communication Channel

The underwater acoustic communication channel is different from the RF channel in three respects: (1) the long multipath delay due to sound refraction and long duration of reverberation from the ocean boundary; (2) the severe signal fading due to time-variable transmission loss; and (3) the high Doppler spread/shift, i.e., the variability and offset of receiver frequency and phase relative to the transmitter resulting from the media and/or platform motion.

Multipath delays in underwater acoustic channels can last tens to hundreds of milliseconds, causing inter-symbol interference to extend over tens to hundreds of symbols depending on the carrier frequency and symbol rate. Inter-symbol interference in RF channels is orders of magnitude less and thus easier to deal with. Doppler shift of carrier frequency in underwater acoustic channels is several orders of magnitude larger than that of the RF channel since the sound speed is many orders lower than the speed of light. Hence, carrier frequency identification and symbol synchronization are critical for underwater systems. In addition, Doppler spread is non-negligible in the underwater communication channel as sound propagates through a random ocean medium and scatters from moving surfaces.

In a random medium, signal phase and amplitude fluctuations resulting from propagation through random environments are range, source, and receiver depth- and frequency-dependent. The temporal scale of fluctuations dictates the rate of adaptation for a coherent processor. The magnitude of the fluctua-

tions determines how well the adaptation will work. Since successful communications require a sufficient input signal-to-noise ratio (SNR), appropriate placement of the source and receiver is necessary to avoid the "shadow" zones (areas where transmission loss is high). Random media increase the probability of signal fading; signal fading occurs when multipath arrivals interfere destructively. One strategy to ameliorate signal fading is to use multiple receivers (spatial diversity). The other is to use array beamforming to improve the SNR.

Phase-Coherent Processor

The phase-coherent processor uses an adaptive channel equalizer to remove the inter-symbol interference and a phase-locked loop (PLL) to compensate for the fast phase fluctuations. The equalizer involves a feed-forward loop and a decision feedback loop (Fig. 1). The equalizer and the PLL are jointly updated to adapt for the fluctuation of the channel. An initial sequence of training data (symbols known exactly by the receiver) is used to train the processor. After training, a decision is made for each symbol by the smallest error between the estimated symbol and the true symbols (e.g., four symbols for QPSK). Symbol synchronization is done by timing from an initial pulse code, such as a Barker code, which can also be used as a wake-up signal for the processor.

The performance of the processor bears on the tap coefficients of the feed-forward and feedback loops. The processor must determine how many coefficients will be used and how fast they will be updated. Currently, user input may be required to achieve optimum performance. For an operational Naval system, it is imperative that user-input "tuning" not be required. This is considered an automation problem. To illustrate the problem, note that the tap coefficients are closely related to the multipath arrivals since their function is to remove/compensate for multipath echo arrivals. Because multipath arrivals are discrete in time (measured in terms of symbols), assigning taps where no multipath arrivals are expected will increase computational error. Also, using unnecessary taps significantly increases processing time, leaving less central processing unit (CPU) power for computation.

On the other hand, too few taps decreases the processor's ability to remove the multipath inter-symbol interference and, consequently, increases the bit error rate (BER). A significant advance in recent years is the development of a multichannel sparse equalizer algorithm that assigns taps where multipath arrivals are expected. Multipath arrivals are determined by estimating the channel impulse response function using the initial pulse (e.g., Barker code) for each channel. In many cases, spatial diversity may be necessary to minimize the BER, hence the need of the multichannel decision feedback equalizer (DFE). Note that channel impulse estimation requires high SNR for each channel. The rate of channel update should be based on some statistical inference measures. The parameters are currently set by experience.

Noting that the channel response and its rate of fluctuation vary from ocean to ocean, it is not readily apparent how the acoustic environmental condition affects the performance of the DFE and PLL; this requires an analysis of the impacting channel physics. Underwater signal fluctuations, unfortunately, do not follow simple statistical distributions; an empirical model will be inadequate. In many cases, there is no clear correlation of the BER with observable envi-

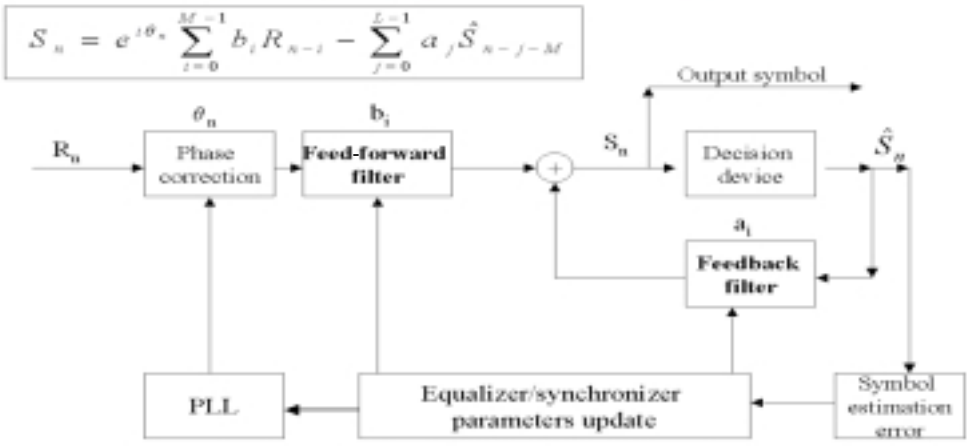


FIGURE 1
Schematic diagram of decision feedback equalizer and phase-locked loop.

ronmental acoustic variables. In environments that produce many diffused multipath arrivals, the equalizer may have to work harder (requiring longer training data) or may fail (producing high BER). But the equalizer may also fail in simple multipath environments when other conditions have changed.

The NRL ACOMMS project was created to investigate how the acoustic environment affects the performance of phase-coherent underwater acoustic communications. The long-term objective of this project is to quantitatively determine how and under what conditions the random processes in the ocean and the ambient background noise field affect the data transmission rate and bit error rate of phase-coherent acoustic communication systems. Such results could be used for predicting underwater acoustic modem performance in operational areas of interest. The short-term objectives are to relate (or correlate) the characteristics of signal propagation with the BER performance of the phase-coherent algorithm and to develop new algorithms to improve the BER. At-sea experiments are an integral part of the project because data are a prerequisite for identifying the issues affecting the processor performance and for validating new algorithms.

PERFORMANCE SIMULATIONS AND PREDICTIONS

Random fluctuations in the ocean originate from complex dynamic oceanographic processes that are not always independent of each other. These processes cause sound speed variation as a function of range, depth, and time. Since oceanographic measurements are undersampled in space and time, sorting out the sound speed variations according to the processes is not possible in practice. For this reason, a simulation model is useful for determining which oceanographic process has the most effect on acoustic communications. Only a handful of processes are expected to play a dominant role; the particular processes that need consideration depend on the acoustic frequency. For midfrequencies (2-5 kHz), surface waves, internal waves, and turbulences are important processes to consider. At high frequencies (10-30 kHz), surfaces waves, bottom roughness, turbulence, and other fine-structure inhomogeneities in the water column dominate acoustic propagation. All these processes have a wide spectrum in wavelength that overlaps the wavelengths of the ACOMMS signals. For ACOMMS analysis, the critical parameters are the temporal scales of the fluctuations; for many processes, the temporal scale can be determined from their wavenumber-frequency dispersion relations. One finds that small surface waves and turbulence cause rapid fluctuations in the phase and amplitude of the ACOMMS signals. Internal waves have a much slower fluctuation rate but can cause severe amplitude fading. Although bottom roughness is stationary, it affects the signal by the fact that each pulse hits a different part of the bottom due to surface and media fluctuation and platform motion. As a result, large surface waves also play a role.

For quantifying ACOMM performance, a simulation model consisting of three parts was created: (1) A physics-based random media model simulates sound speed variations caused by various oceanographic processes. This model generates realizations of sound speed variations as a function of range, depth, and time. (2) A full-field acoustic model propagates broadband impulse response functions through the time-varying, range-dependent sound speed fields. The time-varying impulse response functions are convolved with the transmitted signals to produce the received signals at the receiver array. (3) An ACOMMS performance analysis model evaluates the BER through the time-evolving channel.

The random media model uses parameter values based on previous experimental measurements as much as possible. Although internal waves and turbulence are modeled with different analytic expressions in the literature, we used an "approximate" spectrum that can fit both internal waves and turbulence with a change of parameter values. It is realized that the model may not be perfect, but it captures the basic physics and is expected to yield a first-order understanding of the environmental impact on the ACOMMS performance.

The simulation analysis is focused on two areas: the temporal coherence and phase fluctuation. The effect of medium fluctuation on the acoustic channel can be characterized by the temporal coherence of the impulse response function, which is defined as the normalized value of the correlation of the impulse response function, with the response function at an initial time. Its magnitude lies between 0 and 1. The temporal coherence gives an indication of how much (the coherence value) and how fast (the coherence time) the channel has changed in time. The higher the coherence loss, the harder the equalizer must update. The shorter the coherence time, the faster the equalizer needs to update. Since the equalizer is a numerical algorithm based on the minimal mean-square error criterion, there is a limit in its performance, depending on the algorithm. Using the coherence value as a measure, we find that the recursive least-square (RLS) error algorithm breaks down whenever the coherence value drops below 0.5. This conclusion is based on numerous simulations using both the conventional and sparse equalizer. We found

that the equalizer typically requires a minimum length of time (20 to 30 symbols) to adapt to a changing environment. If the channel changes in less than this minimum time required to update (coherence time < update time), the equalizer may never catch up with the channel.

For phase-modulated signals, fluctuation of the signal phase is a key variable that will impact the algorithm performance. This is because the phase of the signal changes faster than the amplitude. The PLL is implemented to remove the fast phase change that the equalizer is not able to track. The phase change of the signal can be measured by the phase spectrum, which is the Fourier transform of the phase change with respect to time. The impact of the phase change depends on the spectrum value. We found, not surprisingly, that the algorithm can tolerate slow phase changes more easily than fast phase changes. The amount of phase change tolerable by the PLL (used in conjunction with the equalizer) is thus a function of the frequency at which the phase changes. The maximum tolerable phase change is evaluated in the simulation study for a channel with a fixed multipath arrival structure. The result is plotted in Fig. 2 as a threshold curve. A high BER is expected when the channel phase fluctuation spectral values exceed the threshold curve.

Other signal properties must be included to simulate the ACOMMS performance. These include the statistical distribution of the amplitude, the spatial coherence of the signal, and the change of transmission loss caused by the random media. We used values estimated from data wherever applicable.

RECENT ADVANCES IN PHASE-COHERENT COMMUNICATIONS

The state-of-the-art multichannel sparse decision-feedback equalizer has been tested in the ACOMMS ATD program at both mid and high frequencies. The results showed that a substantial percentage of the packets were error-free with error encoding over a wide range of distances. For QPSK signals, bit-error-free packets require ≥ 12 dB SNR at the input level; multiple channels (3-4) are required for the equalizer to work; and the performance is occasionally poor due to poor channel properties and poor ability to estimate the channel impulse response. These results are common to other tests as well. Many tests have shown that BER can often be improved in post analysis by playing with the number of taps and update parameters; BER is usually high when using only one channel; and a 5-s limit on the communication packet length was used in practice at midfrequency, with duty cycle less than 50%. This means that the average data rate is less than half of the burst data rate.

NRL has developed several new algorithms to improve the performance of the multichannel DFE at midfrequencies. The new algorithms extended phase-coherent acoustic communications to low SNR (~ 1-2 dB) input cases, extended the communication packet length from 5 seconds to 10-20 seconds, and

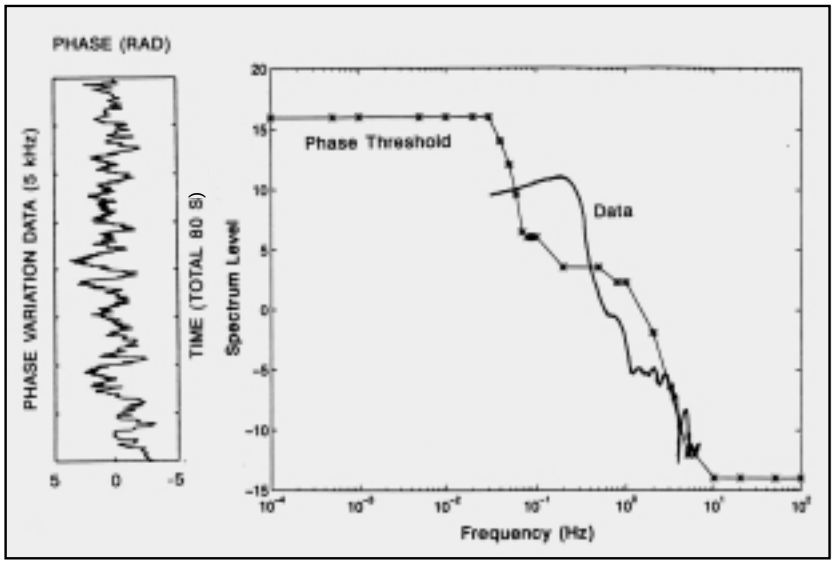


FIGURE 2
Example of phase variation time series (left) and its spectrum (right). The phase threshold (right) indicates the maximum spectral level of phase fluctuations that can be tolerated by the phase-locking loop.

reduced the number of channels from 4 to 1-2. The improved performance has been demonstrated with at-sea data as discussed below; more at-sea tests are being planned. The new algorithms have also been demonstrated to work with one set of parameters for several different environmental conditions. These accomplishments pave the way for: potential covert communications, which is a high priority for Navy operations; more energy efficient modems due to the use of longer packet length and less computational load (less number of channels); and modem algorithms capable of working autonomously in different oceans.

The improved algorithms are based on inferences drawn from physical analysis of the temporal and spatial coherence, as discussed above. Temporal fluctuation, particularly the phase fluctuation, determines the BER performance and the length of the packet. It was observed that large phase errors resulting from incorrect Doppler shift estimation could exceed the threshold values for the PLL. The error in the carrier frequency estimation could ruin symbol synchronization for a long communication packet. During the RDS2 experiment, communication packets of 10-s length were transmitted using a towed source to a drifting vertical array. QPSK signals with a 3.5 kHz carrier frequency and 500 Hz bandwidth were used. Four channels were required to process the packet using the existing algorithm, and the results were good only for the first ~5 seconds (Fig. 3). A continuous-wave (CW) tone was transmitted before the QPSK packet. Using the CW signal to estimate the correct carrier frequency and then correct the synchronization error, it was demonstrated that the entire 10-s packet was successfully equalized using only one channel (Fig. 4). This result was consistently obtained for all the 10-s packets that were transmitted during the 2-hour run in the RDS2 experiment. We note that by removing the phase due to Doppler, the residual phase is apparently small enough to fall below the phase threshold curve. The actual phase spectrum is not easily deducible from the data.

Another improvement comes from the study and use of spatial coherence to make the equalizer work with low-input (~1-2 dB) SNR. This was demon-

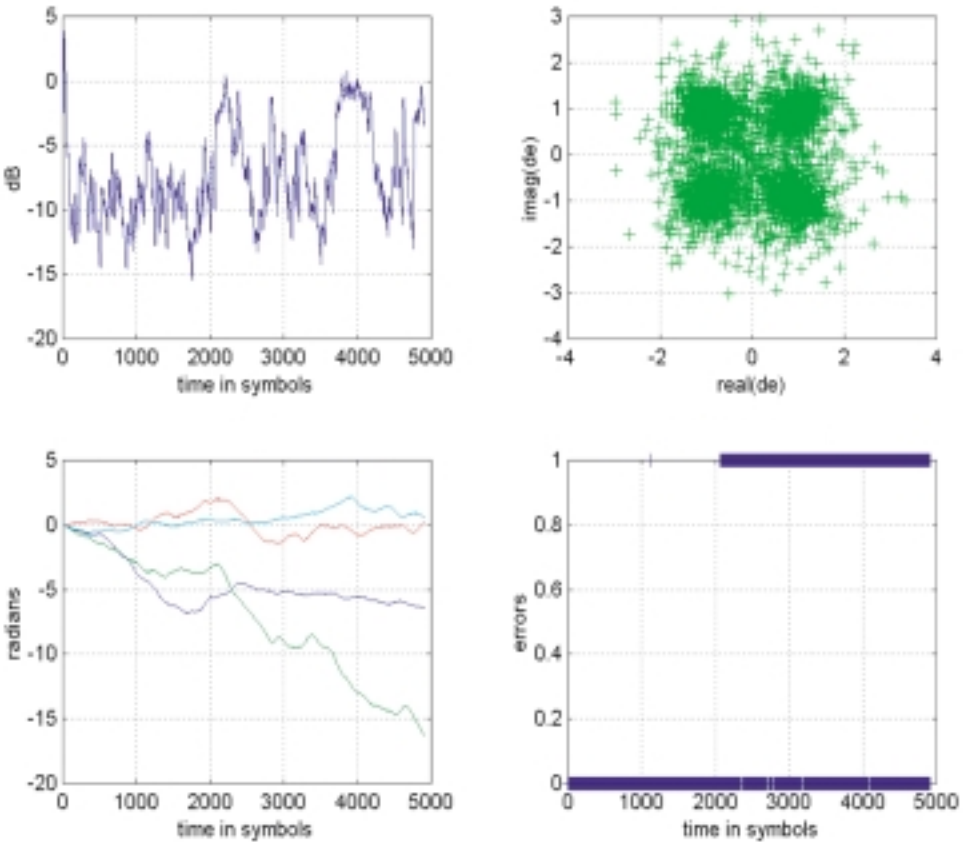


FIGURE 3
Plots of mean-squared-error (upper left), symbol constellation (upper right), phase compensation (lower left), and bit-error (lower right) versus the symbol number for a 10-s long QPSK signal. The curves in the lower left figure correspond to different channels.

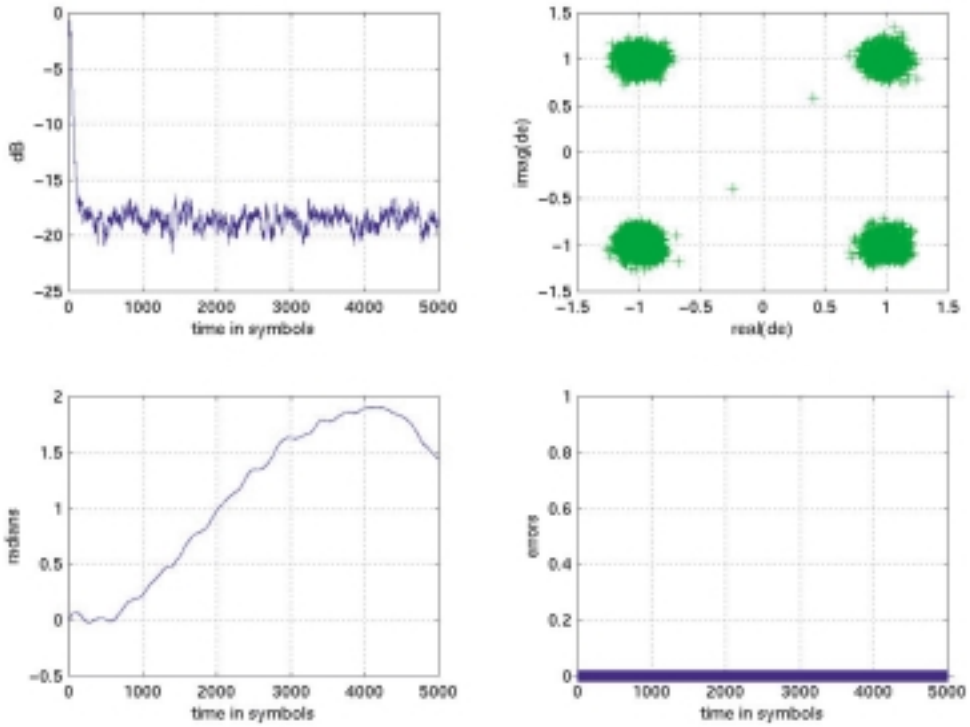


FIGURE 4
Same as Fig. 3, processed using the correct Doppler shift.

strated using the ACOMMS data collected during the ADVENT 99 experiment and concerns the use of multichannel DFE. Note that multichannel receivers (spatial diversity) were originally used to combat signal fading. In this scenario, the multichannel signals must be uncorrelated (spatially) so that they do not fade at the same time. Multichannel DFE determines the relative phase and weighting of the channel feedforward coefficients based on the minimum square error criterion. It requires high (~ 10-15 dB) SNR. Likewise, estimation of the channel impulse response requires high SNR data. For the multichannel DFE to work with low SNR data, additional processing gain and/or array gain will be required to enhance the SNR before applying the multichannel DFE. Processing gain can be achieved by spreading the signal in time using, for example, a code division spread spectrum (CDSS) scheme. Array gain can be achieved by beamforming an array of receivers. However, we note that beamforming requires the signal to be coherent over the receivers; this requirement is opposite to that of spatial diversity.

The solution is to use a combination of beamforming with spatial diversity. Subarrays of closely spaced phones are used to provide required array gain. To use multichannel DFE, the centers of the subarrays are widely spaced so that the beamformed array outputs are spatially uncorrelated. This requires a selection of phones based on the spatial coherence. We did such a study using the ADVENT 99 data. The ADVENT 99 experiment was conducted in the Adventure Bank, Sicily, in 1999 with the support of the NATO SACLANT Undersea Research Center. A fixed source and receiver array were used so that Doppler shift would not be an issue. QPSK signals with a 1.2 kHz carrier frequency and 500 Hz bandwidth was transmitted to a range of 10 km. The source was mounted at 2-3 m above the seafloor. An array of ~20 phones was used to investigate the performance of multichannel DFE (spatial diversity) versus the phone separation and number of phones. An array of 10 phones spaced 2 m apart was used to compare the performance of the combined beamforming/spatial diversity method with that using only the multichannel DFE. Figure 5 shows the BER of these two methods. The multichannel DFE used 4 phones separated by 6 meters so that the signals were not correlated. (The correlation length was ~ 4 m, as determined from the data.) For the combined beamforming/spatial diversity method, we used three overlapping subarrays, each containing five phones. The 4-channel DFE worked when the SNR was above ~ 8-10 dB but failed (BER >80%) for SNR < 8-10 dB. The NRL method worked well, even when the SNR at the phone level was as low as 1-2 dB. Figure 6(a) shows the channel impulse esti-

mated from the Barker code for the ~2 dB SNR case for which many (unnecessary) tap coefficients will be selected based on a threshold. Figure 6(b) shows the impulse response estimated after beamforming for which the tap assignment will be much less ambiguous. Comparing Fig. 6(a) with Fig. 6(b) one finds an array gain of ~ 6-7 dB as a result of beamforming. We have modified the tap assignment program to eliminate the taps that represent low-level multipath arrivals or noise fluctuations.

The performance prediction for the temporal coherence was tested during the Littoral Warfare Advanced Development (LWAD) 98-1 experiment in the Gulf of Mexico area. Time diversity was used to process data collected on a sonobuoy. The data have SNR > 10dB, so SNR is not a likely cause for BER. It was found that large BER were associated with data with signal coherence < 0.5. This result supports the performance predictions based on temporal coherence discussed in the previous section.

SUMMARY AND FUTURE WORK

Underwater connectivity is critical for many Naval operations and is an integral part of future net-centric warfare activity. Underwater connectivity requires a high-data-rate wireless acoustic communications link between surveillance sensors, UUVs, submarines, and surface ships for command, control, and data telemetry. Recent advances in phase-coherent communications technology make possible high-data-rate communications at long ranges. Real-time operating high-data-rate modems are being developed that could work reliably under diverse ocean conditions and locations. Continued ONR/NRL research can be expected to resolve environmental influences on the modem performance

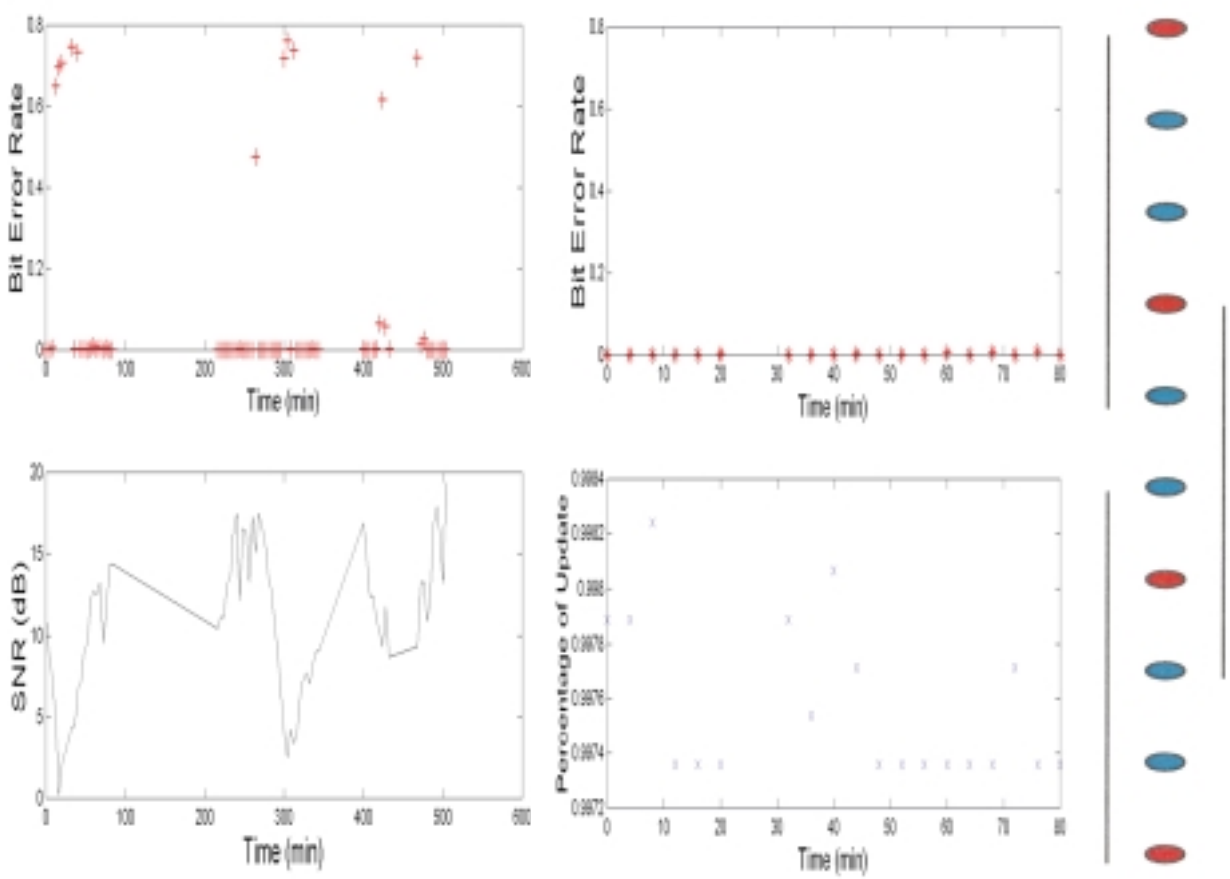
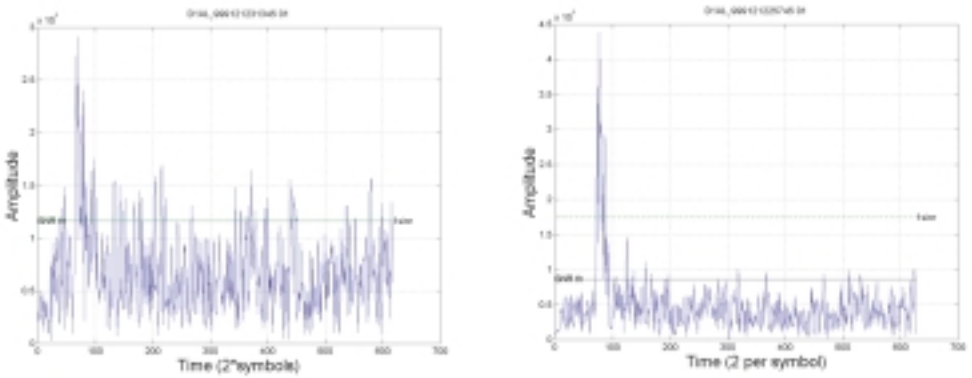


FIGURE 5
BER using 4-channel DFE (upper left) and the corresponding input SNR (lower left) for a period of ~ 9 hours. The upper right figure shows the BER using the combination of beamforming with spatial diversity for the first 80 min of the data. The BER was minimal, even when the SNR was as low as ~1-2 dB. The lower right figure shows the rate of update of the equalizer. The far right figure shows the array configurations. The red dots indicates the phones used in the multi-channel DFE. The arrows indicates the subarrays used for beamforming as discussed in the text.



(a) (b)

FIGURE 6
Left figure shows the impulse response for a single channel with low SNR; right figure shows the same after beamforming using five sensors.

and develop improved algorithms for covert communications, multi-access networking, etc.

The NRL ACOMMS projects followed a physics-based approach that focused on identifying critical signal properties that have the most impact on the BER. Three initial performance predictions were made based on the temporal coherence length, value, and phase fluctuation spectrum. This analysis laid the framework for an improved algorithm that allows longer communication packets (10-20 s at mid frequencies) and requires fewer channels in the decision equalizer. The study of spatial coherence led to a method for combining beamforming with spatial diversity. The new method has been demonstrated to work with some data having ~1-2 dB input SNR; further tests of this new method are being planned. The low SNR requirement will extend the communication range and make covert communications possible.

Assuming the success of a local area underwater acoustic network with multi-user access, distributed networks of autonomous sensor nodes with acoustic modems could be set up in many parts of the ocean. The autonomous sensor nodes could be affordably mass-produced. Independent nodes offer the important benefit of architectural flexibility, permitting sensor-field aperture and spacing to be optimized for the intended mission. These networks could serve the dual purpose of civilian weather and oceanographic research on one hand, and acoustic ocean surveillance for the Navy on the other. To the user, the acoustic local area network will not be any different than the standard network; it will use modem handshaking protocols already developed in the RF and Internet world. As such, it is expected that the creation of an "oceanet" could greatly expand the capability for weather nowcasting and forecasting, marine mammal and fishing migration, and many other areas of undersea research. Developing an asynchronous multiple-user acoustic communications network, therefore, will be an important research topic for the future.

ACKNOWLEDGMENTS

The NRL ACOMMS experiments were the results of a collaborative effort of many people including particularly, A. Al-Kurd, E. Carey, P. Gendron, M. McCord, and J. Schindall. The support of LWAD, TTCP RDS program, and NATO SACLANT Center for Undersea Research were greatly appreciated; without this, most of the ACOMMS experiments could not have been possible. J. Schindall has developed most of the software for at-sea data transmission/collection and initial data reduction. A. Al-Kurd has conducted extensive acoustic simulations and processed data from several experiments. K. Yoo has carried out acoustic signal propagation modeling in random media to support ACOMMS data analysis and interpretations. The support, encouragement, and leadership provided by our branch head, M. Orr, are indispensable to the ACOMMS project. Comments on the manuscripts by P. Gendron, P. Mignerey, and M. Orr are greatly appreciated.

[Sponsored by ONR]

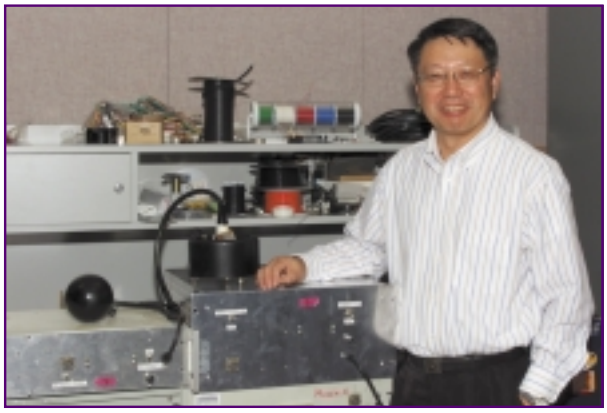
REFERENCES

There are too many references to be cited here. For recent NRL accomplishments, see the following references.

[1] T.C. Yang and A. Al-Kurd, "Performance Limitations of Joint Adaptive Channel Equalizer and Phase Locking Loop in Random Oceans: Initial Test with Data," *Proceeding of OCEANS '00*, Vol. 2, 803-808, 2000.

[2] T.C. Yang and M. Siderius, "Phase Coherent Communications at Low Frequencies During the Advent 99 Experiment in the Sicily Strait," *Proceeding of OCEANS '00*, Vol. 2, 1005-1010, 2000.

THE AUTHOR



TSIH C. YANG came to NRL in 1979 and worked on Arctic acoustics until 1993. He discovered ice edge noise hot spots, published many new findings on under-ice reflectivity, low-frequency sound transmission loss, scattering and reverberation, ambient noise coherence and directionality, ice floe vibrations, and geophone array response. He was responsible for the array wet end design for the Navy Ice Mounted Surveillance System. Since 1993, he has worked on geo-acoustic inversion of ocean bottom properties, sonar array processing, and underwater acoustic communications. He pioneered matched mode processing for a vertical line array and recently developed matched-beam processing for Navy horizontal line array systems. He formerly was the head of the Arctic Section and now heads the Dispersive Wave Guide Effects Group. He received three patents and one Alan Berman Research Publication Award. Dr. Yang is a fellow of the Acoustical Society of America.